



Hybrid information flow analysis against web tracking (invited talk)

Thomas Jensen

► To cite this version:

Thomas Jensen. Hybrid information flow analysis against web tracking (invited talk). CRiSIS 2017 - 12th International Conference on Risks and Security of Internet and Systems, Sep 2017, Dinard, France. pp.1-33. hal-01658896

HAL Id: hal-01658896

<https://inria.hal.science/hal-01658896>

Submitted on 3 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Joint work with:



Frédéric Besson



Nataliia Bielova

Hybrid information flow analysis against web tracking

Thomas Jensen, Inria

Keynote, CRISIS'2017

Overview

1. Web tracking.
2. Knowledge-based information flow analysis.
3. Browser randomisation.

Thanks to EU ePrivacy Directive



Cookies on the BBC website

We use cookies to ensure that we give you the best experience on our website. We also use cookies to ensure we show you advertising that is relevant to you. If you continue without changing your settings, we'll assume that you are happy to accept all cookies on the BBC website. However, if you would like to control your cookie settings at any time.

bcc.co.uk

Continue
Find out more

BBC News Sport Weather Travel Culture Autos TV Radio More... Search

emp.bcci.co.uk **NEWS** 2 June 2013 Last updated at 20:00 GMT **googleads.g.doubleclick.net**

Home UK Africa Asia Europe Latin America Mid-East US & Canada Business Health Sci/Environment Tech Entertainment Video

Magazine In Pictures Also in the News Editors' Blog Have Your Say World News TV World Service Radio Special Reports

LATEST: South African officials investigate claims that Muammar Gaddafi and his family stashed \$1bn in assets in the country

Protesters return to Turkey streets

Hundreds of protesters return to the streets of Istanbul and Ankara, with the PM accusing some elements of trying to undermine democracy.

892

Determined to stay
Media slams handling of protests
Is Turkey's secular system in danger?
In pictures: Saturday clashes

effectivemeasure.net

b.voicefive.com

Magazine

Watching brief

pagead2.googlesyndication.com

Eden's marshes
Restoring the wetlands drained by Saddam

Features

js.revsci.net

The Queen's 'dazzling' coronation - 60 years on

Syrian rebels and Hezbollah 'clash'

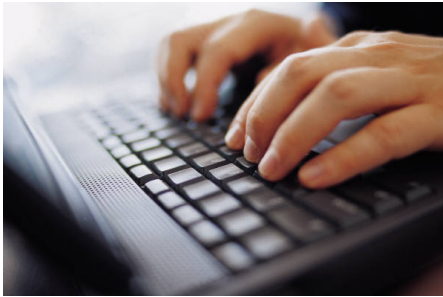
A number of people are killed in rare clashes on Lebanese soil between Syrian rebels and the Lebanese militant group Hezbollah, say reports.

Qusair's strategic importance
Hezbollah's role
Syria town
Unwinnable war

googletagservices.com

b.scorecardresearch.com

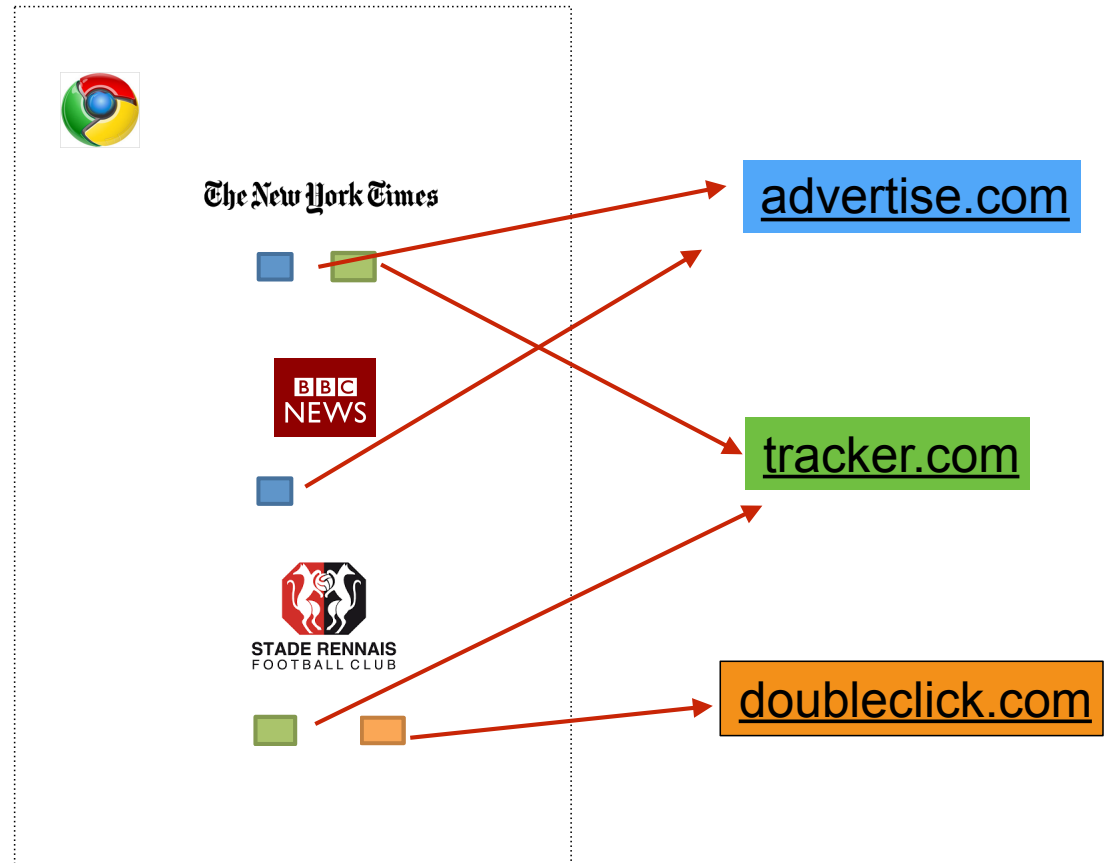
Web tracking



Bigger browsing profiles

= increased value for trackers

= reduced privacy for users



(Hypothetical tracking relationships only.)

Browser extensions to the rescue?



AdBlockPlus: blocks scripts/requests **only from known advertisement companies**

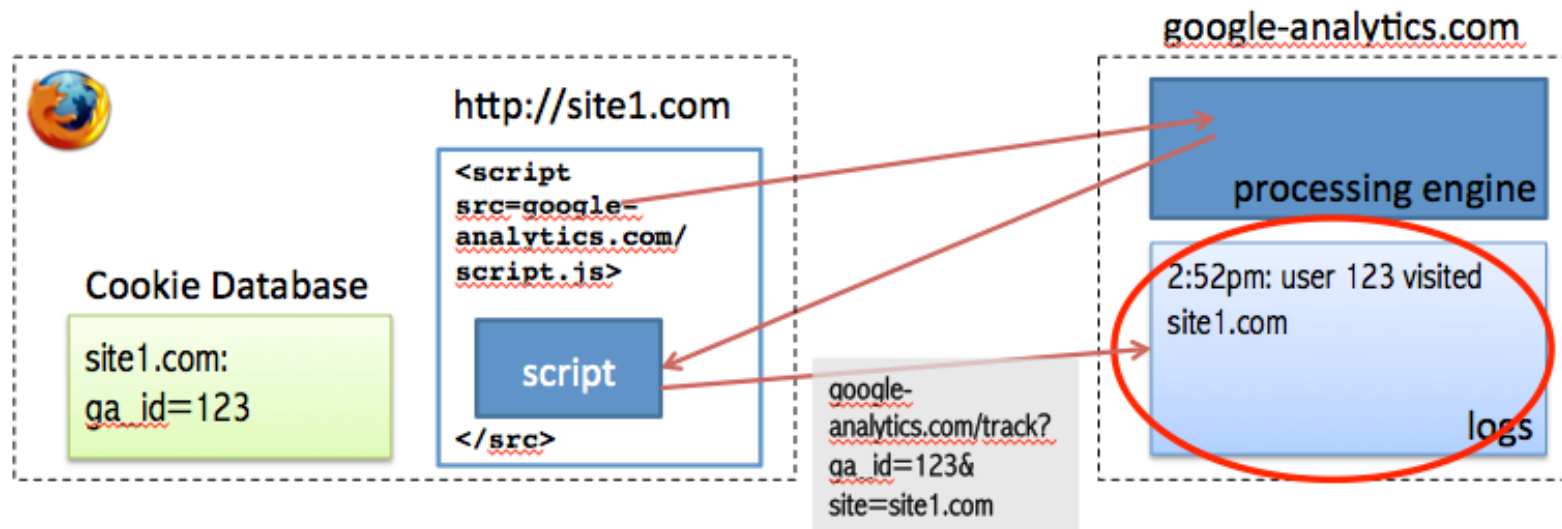


Ghostery: blocks scripts/requests **only from known tracking companies**

- No protection from tracking by other companies
- No protection from tracking by the main (first-party) web site

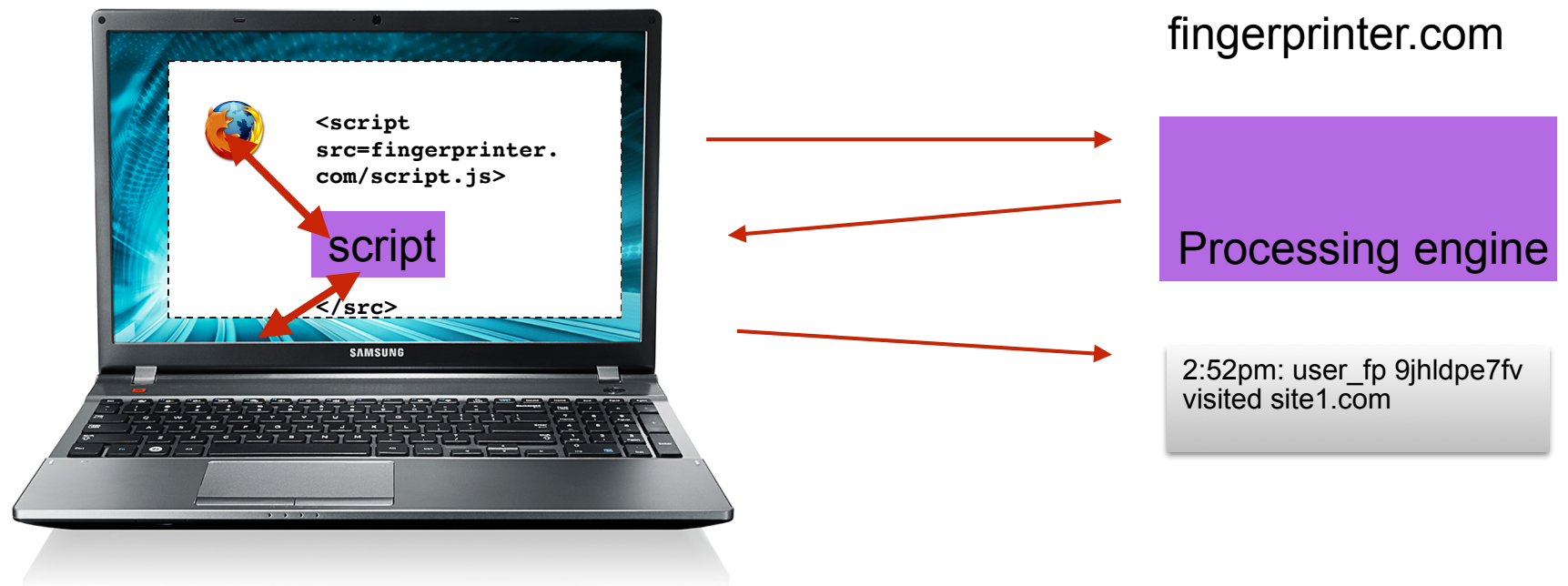
Tracking by storing identity

Cookies are used to track repeated visits to a site.

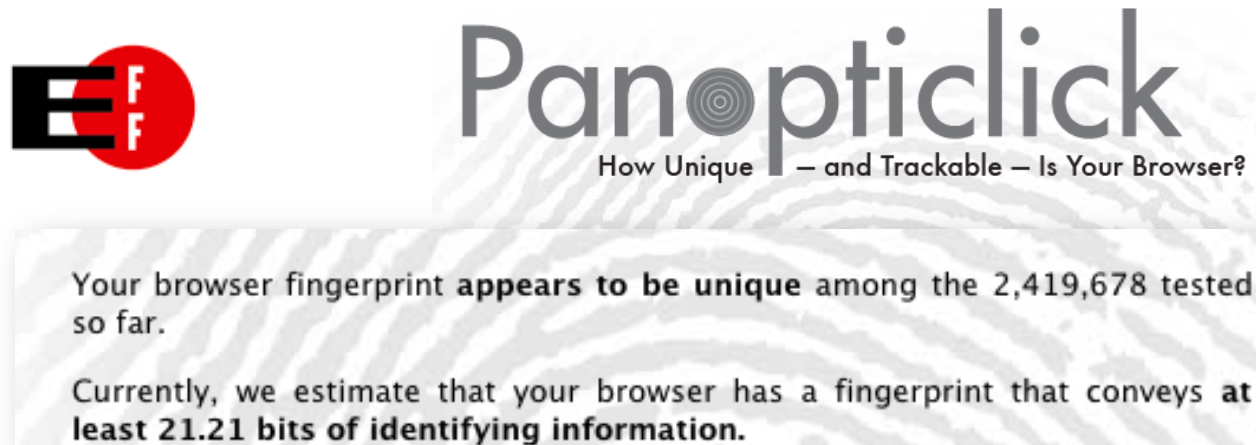


Tracking by finger-printing

Browser and operating system properties are used to track repeated visits to a site.



Panopticlick: an early study on fingerprinting



- A study from 2010 showed that 83,6 % of browser fingerprints were unique among the 500 000 browsers tested.
- **Fingerprints:** HTTP headers, browser and OS features, language, plug-ins, fonts, ...

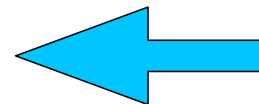
Protection mechanisms

Stateful tracking: well-known and being addressed at different levels:

- Third-party cookie blocking
- EU e-Privacy directive
- Non-interference analysis for JavaScript

Stateless tracking: less clear

- IP address tracking
- Web browser finger-printing



2. Knowledge-based information flow analysis

Use information flow to spot web tracking

Scripts can access information about configurations

Goal: Estimate the information that a script learns about configurations.

But some information leakage is reasonable!

Scripts may need information about browser and execution environment for optimal user experience.

Information flow control

Non-interference

Secret input does not flow into public output

Yes/no ?

Attacker knowledge

What information about the secret flows to output in an execution.

Secret > 42

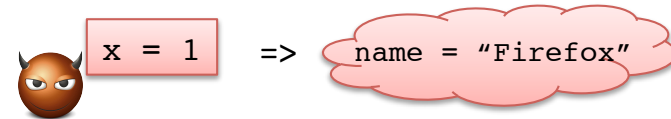
Quantitative information flow

How much information about secrets is leaked?

Leaks 17 bits

Attacker knowledge

```
var x = 0;
if (name == "Firefox") {
    x = 1;
}
else {
    if (fonts == fontsSet1) {
        x = 2;
    }
}
output x;
```



Depending on the browser, **different executions** leak **different amount of information**.

Non-interference

A program is termination-insensitive non-interferent if it produces the same output for all low-equivalent input.



Example with two binary secrets $h1, h2 \in \{0,1\}$

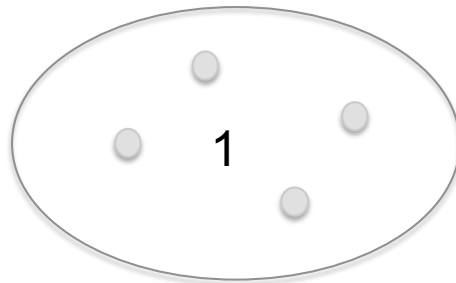
Secure and insecure executions

The following program is not secure - but it has secure executions.

```
if h1 = 1 then x = 1
else skip;
if h2 = 1 then low = 1
else low = x;
output low
```

Initial values:

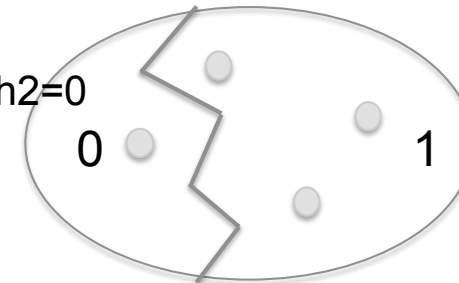
[low = 1, x = 1]



Secure

[low = 1, x = 0]

$h1=0 \wedge h2=0$



Leaks information about secret


Monitoring information flow

Dynamic information flow control

- Label values with security labels.
- Propagate labels during execution.
- Halt execution before outputting secret values.

Non-interference needs "no-sensitive- upgrade" principle:
no assignment to public variables under secret control

```
[secret = 1]
```

```
if secret = 1 then  
 output = 1  
else output = public;  
return output
```

Hybrid information flow analysis

Combine dynamic IF control with a static analysis

- statically analyse the non-executed branches
- admits more executions but limited by the precision of the static analysis

```
[secret = 1, public = 1]
```

```
if secret = 1 then  
  output = 1  
else output = public;  
return output
```

Static analysis says
public : [1]



Hybrid monitor of attacker knowledge

Define a hybrid monitor that tracks the knowledge of an attacker.

Use knowledge to

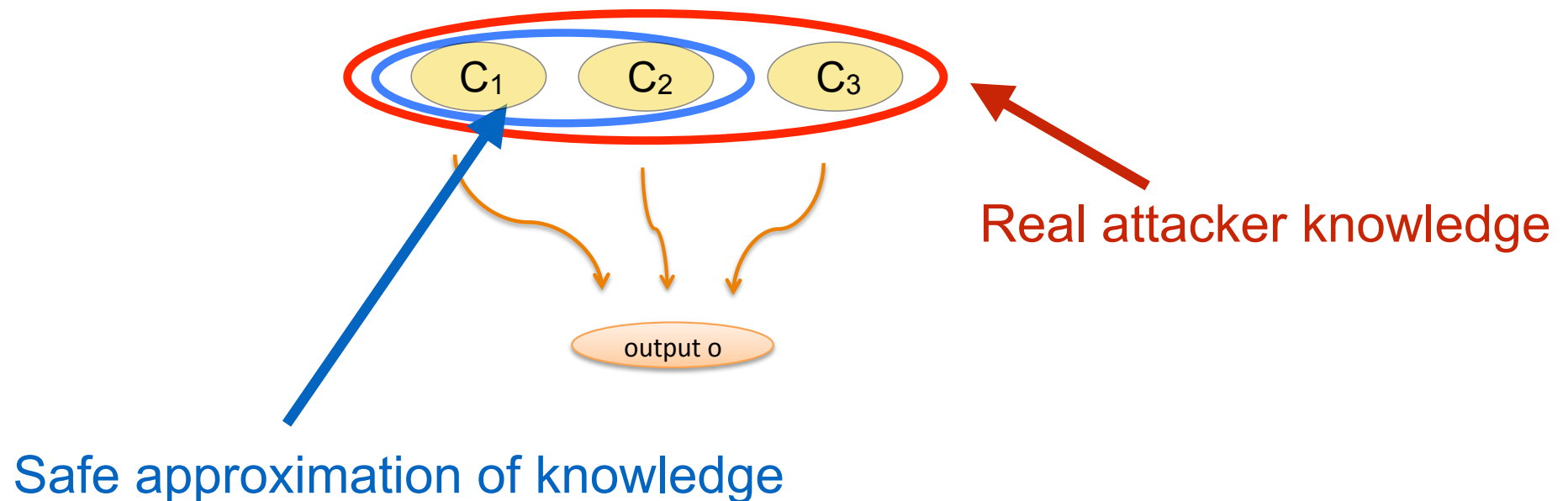
- enforce non-interference,
- accept more secure executions.

Solution:

- A domain for representing attacker knowledge.
- Monitor formalised as an operational semantics.

Attacker knowledge

Knowledge: What does an attacker learn about the input state by seeing a particular output.



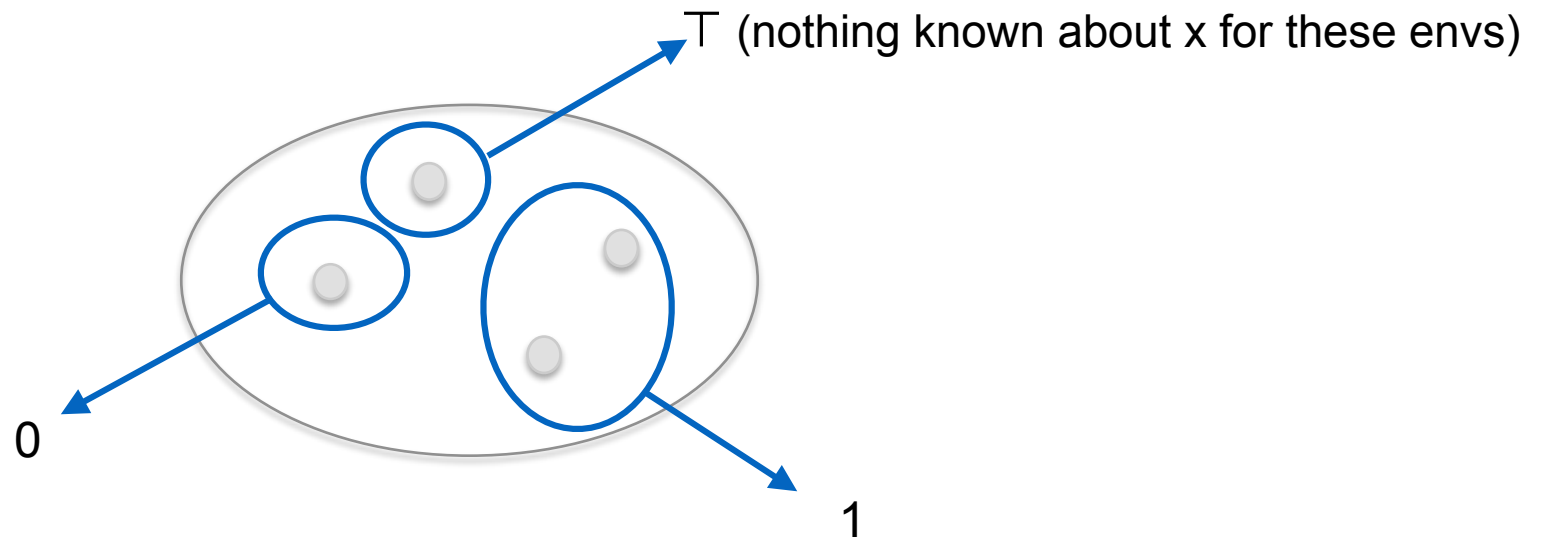
Domain of knowledge

$$\kappa: \text{Var} \rightarrow \mathbf{K}$$

$$\mathbf{K} \triangleq \text{Env} \rightarrow \text{Value} \cup \{\top, \perp\}$$

The knowledge in a variable $\kappa(x)$:

which initial environments leads to a particular value of x ?



The hybrid monitor

Defined as an operational reduction semantics:

$$(P, \rho, \kappa) \Downarrow (\rho', \kappa')$$

- $\rho, \rho' : \text{Env} \cup \{\cdot\}$,
- $\kappa, \kappa' : \text{Var} \rightarrow \mathbf{K}$ labeling of variables with knowledge,
- $\text{Env} : \text{Var} \rightarrow \text{Val}$ for dynamic analysis,
- \cdot "empty" environment for static analysis.

Semantic rules

Rules come in pairs: one dynamic and one static

Dynamic analysis computes real values and knowledge

Static analysis called on-the-fly and propagates only knowledge

$$\text{IFDYN} \frac{C[e]_{\rho} = \alpha \quad (c_{\alpha}, (\rho, \kappa)) \Downarrow (\rho', \kappa_{\alpha}) \quad (c_{\bar{\alpha}}, (\cdot, \kappa)) \Downarrow (\cdot, \kappa_{\bar{\alpha}})}{(\text{if } e \text{ then } c_{tt} \text{ else } c_{ff}, (\rho, \kappa)) \Downarrow (\rho', \text{IFF}(\llbracket e \rrbracket_{\kappa}, \kappa_{tt}, \kappa_{ff}))}$$

$$\text{IFSTAT} \frac{(c_{tt}, (\cdot, \kappa)) \Downarrow (\cdot, \kappa_{tt}) \quad (c_{ff}, (\cdot, \kappa)) \Downarrow (\cdot, \kappa_{ff})}{(\text{if } e \text{ then } c_{tt} \text{ else } c_{ff}, (\cdot, \kappa)) \Downarrow (\cdot, \text{IFF}(\llbracket e \rrbracket_{\kappa}, \kappa_{tt}, \kappa_{ff}))}$$

Operator combining knowledge obtained in branches with knowledge from the condition

Enforcing non-interference

Before any output, the monitor checks non-interference:

- if non-interference then output is not modified.

$$\text{OUTNI} \frac{(c, \text{init}(\rho)) \Downarrow (\rho', \kappa) \quad \rho'(x) = v \quad \text{NI}(\rho, \kappa(x), v)}{(c; \text{output } x, \rho) \Downarrow v}$$

The non-interference check is defined as:

$$\text{NI}(\rho, K, v) \triangleq [\rho]_{\perp} \subseteq K^{-1}(v) \cup K^{-1}(\perp).$$

set of low-equivalent env's

Example analysis

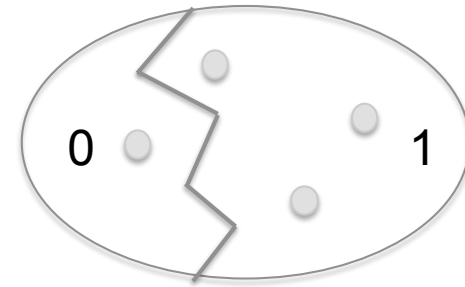
[h1 = 0 , h2 = 1]

Static analysis

```
if h1 = 1 then x = 1
else skip:
if h2 = 1 then low = 1
else low = x;
output low
```

Knowledge

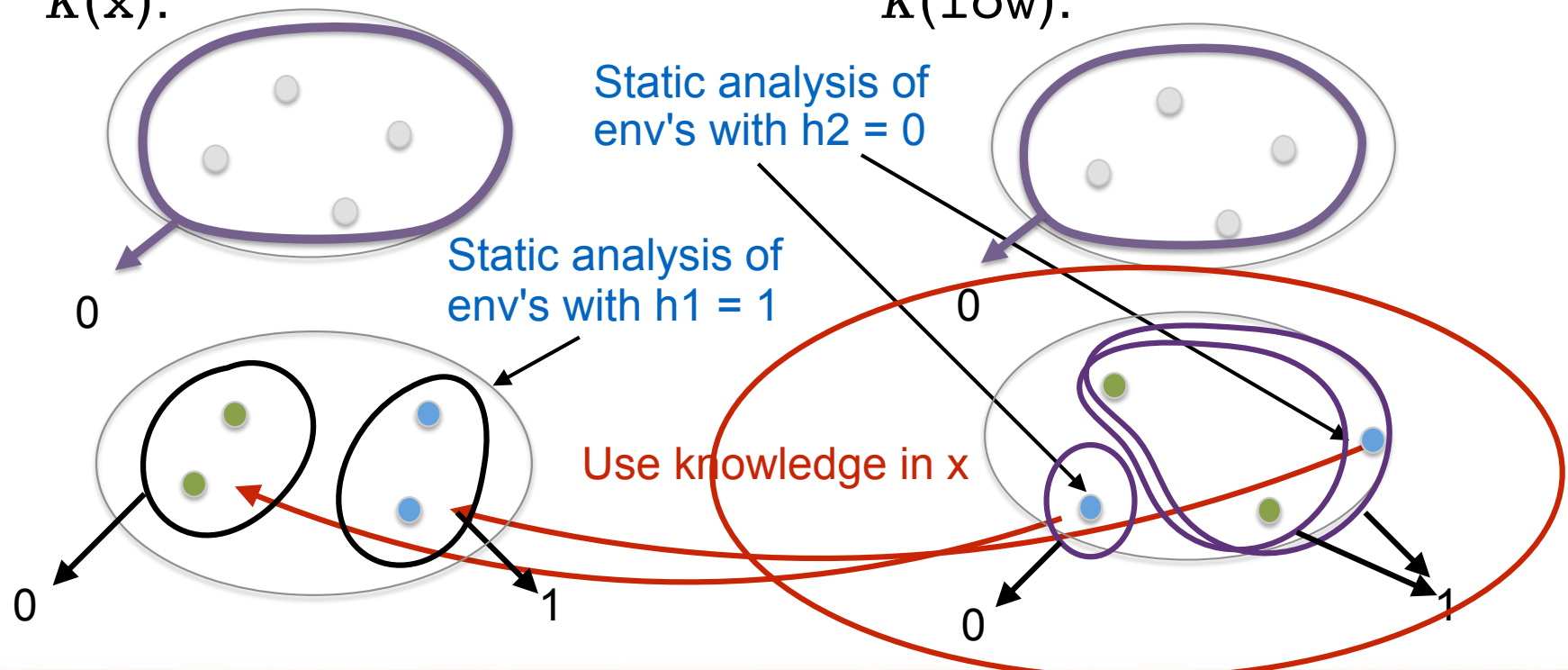
[low = 0 , x = 0]



$K(x)$:

$K(low)$:

Initial



Stock-taking

Knowledge-based information flow analysis

- computes a correct approximation of attacker knowledge,
- enforces non-interference,
- more permissive than existing hybrid monitors,
- can be combined with other monitors to enhance precision.

3. Browser randomisation against web tracking

How to enforce privacy?

The hybrid monitor monitor evaluates how much a tracker learns about an individual user.

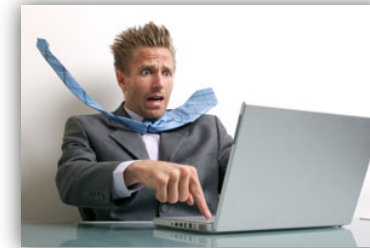
How to react when a script tracks too much?

p(name) =	
Firefox	0.45
Chrome	0.45
Opera	0.10

- With 10 users, our Opera user will be uniquely identified.
- Halting the program or suppress output might still make you identifiable.
- Need collaboration between users.

A solution

- Users switch between configurations
- How many configurations are needed to guarantee privacy?
- Usability issue: users want to switch as little as possible



Challenge: find a distribution on configurations for all users such that

- Privacy is guaranteed
- Usability is maximised

Privacy: probability of guessing

```
if (name == "Opera") x = A;  
else x = B;  
output x;
```

A priori

p(name) =	
Firefox	0.45
Chrome	0.45
Opera	0.10

$x = A$

A posteriori

p(name) =	
Firefox	0
Chrome	0
Opera	1

Probability Pr^G : guessing the secret in case of worst observation.

Stronger that definition based on Bayesian risk:

$$p(A) \max_i p(i|A) + p(B) \max_i p(i|B) = 0.55$$

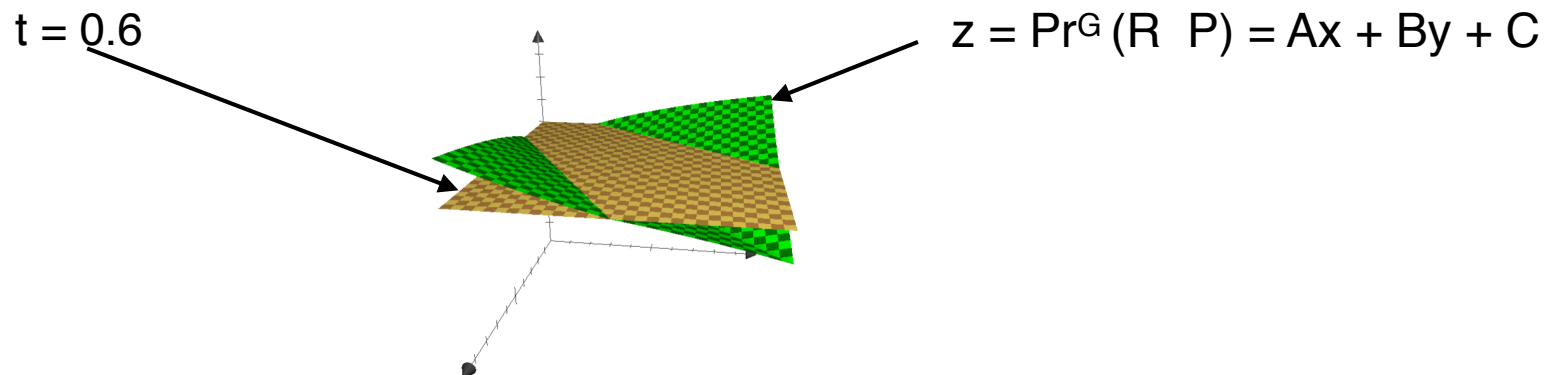
Achieving privacy

Find a randomisation R of user state such that when composed with tracking script P

$$\Pr^G (R ; P) < \text{privacy threshold.}$$

R	"Firefox"	"Opera"
id_1	x	$1-x$
id_2	$1-y$	y

P	o_1	o_2	o_3
"Firefox"	$1/2$	$1/3$	$1/6$
"Opera"	$1/6$	$1/2$	$1/3$



Adding usability

Users don't want to switch configuration:

- choose randomisation so that they get their original configuration as much as possible
- linear constraint on the randomisation matrix.
- problem reduces to linear programming.

R	"Firefox"	"Opera"
id ₁	x	1-x
id ₂	1-y	y

$$\begin{array}{l} \max (x + y) \text{ s.t.} \\ 0 \leq x \leq 1 \\ 0 \leq y \leq 1 \\ Ax + By + C \leq t \end{array}$$

Usability

Privacy

Wrapping up

Web tracking is done by different techniques

- cookies, other browser storage, fingerprinting

Information flow analysis against fingerprinting

- Hybrid information flow monitoring with static analysis
- Computes tracker knowledge.

Enforcing browser anonymity

- Randomisation of configurations
- Combining privacy and usability

Based on a true story...

- F. Besson, N. Bielova, T. Jensen: Hybrid Information Flow monitoring against web tracking, CSF 2013.
- F. Besson, N. Bielova, T. Jensen: Browser randomization against fingerprinting: A Quantitative information flow approach, NordSec 2014.
- F. Besson, N. Bielova, T. Jensen: Hybrid monitoring of attacker knowledge, CSF 2016.